Workshop on Raw Image Formats/May 22, 2008
MEDSBIO

Through the Looking Glass: creating an HDF data prism

Matthew Dougherty
National Center for Macromolecular Imaging/Baylor College of Medicine

A proposal of a pan-image strategy is presented for discussion. A prototype has been created that mounts an HDF file as a virtual file system using FUSE (File system in UserSpacE). This file system in its simplest form is able to intake files and present files transparently to any application program using standard operating system protocols and libraries, such as drag-and-drop and STDIO. Most importantly, more sophisticated possibilities exist that can finesse the image compatibility problem. By using HDF as a virtual file system, transparent conversions across formats are possible and are highly efficient. Most scientific image data formats are file based, and are generally incompatible. Image formats usually consist of two components: metadata that provides entrée into the image, and a contiguous block of pixels. Generally this block of pixels is very simple consisting of one to four dimensions and a simple pixel model such as floating point or integer values. The use of HDF in combination with FUSE will allow images to be written in one format and transparently presented in another image format. By such a method, 99%+ of the image bytes, which are the pixels, can be written one time only, requiring appropriate metadata headers to be created as needed and thus obviating the common practice to duplicate pixels in mass storage. The proposal will also discuss possible avenues of adoption and possible organizational support.