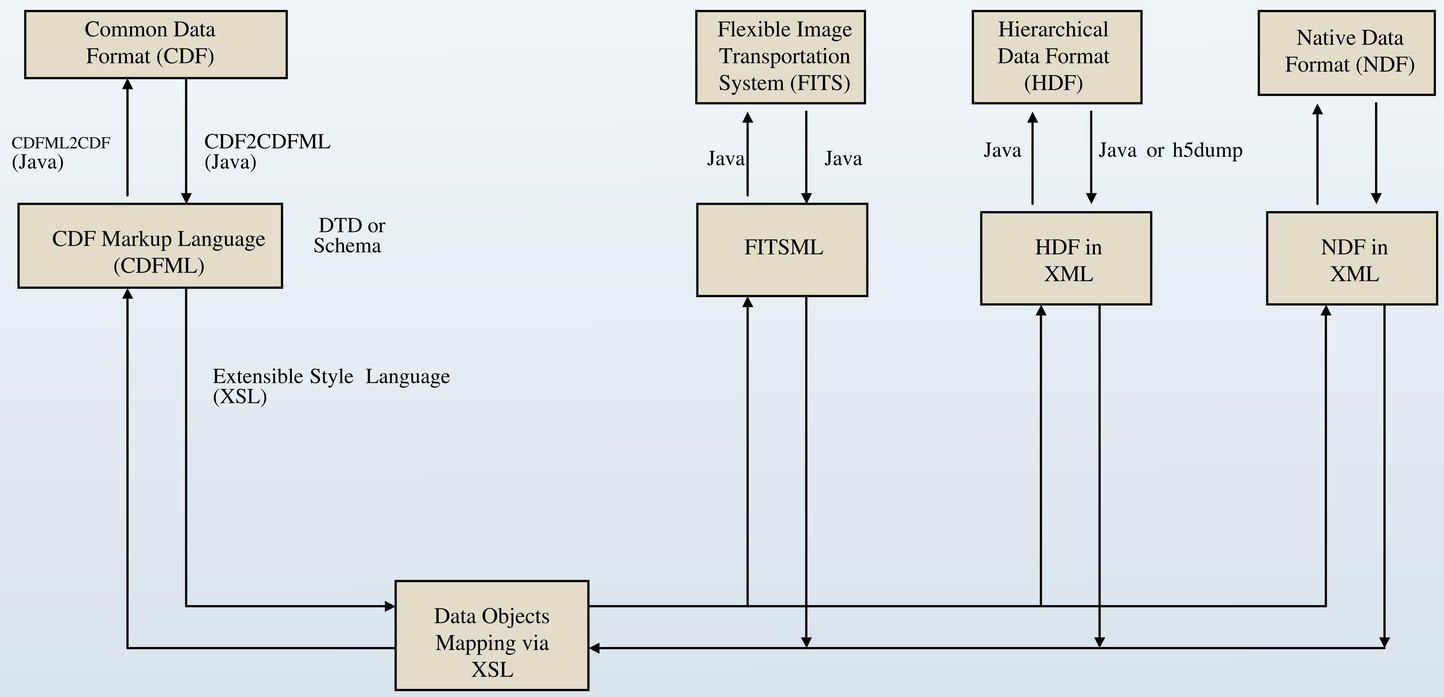
Common Data Format: New XML and Conversion Tools

David Han
GSFC, National Space Science Data Center

Interoperability with Other Data Formats using XML



Abstract

The variety of available data formats (e.g. CDF, netCDF, HDF, etc.) has been a problem for scientists (because data of their interest must be translated into the format they understand before they can analyze data), and it will continue to be a problem for years to come. In a bid to make data format differences transparent to the end users, the CDF office has employed the eXtensible Markup Language (XML) technology and has been developing custom ad-hoc translators to facilitate and promote data interoperability with other data formats.

Software for CDF

- IDL, MATLAB, IBM DX, Application Visualization System
- NSSDC developed Web-based systems
  - CDAWeb (http://cdaweb.gsfc.nasa.gov/cdaweb)
  - SSCWeb (http://sscweb.gsfc.nasa.gov)
  - COHOWeb (http://nssdc.gsfc.nasa.gov/cohoweb/cw.html)
  - OMNIWeb (http://cdaweb.gsfc.nasa.gov/omniweb/ow.html)
- CDAWlib (http://spdf.gsfc.nasa.gov/CDAWlib.html)
  - A set of IDL routines that allow users to manipulate data store in CDF
  - Available at ftp://cdaweb.gsfc.nasa.gov/pub/CDAWlib
- Standard CDF tools (CDFcompare, CDFconvert, CDFedit, CDFexport, CDFstats, SkeletonTable, SKeletonCDF)
  - Distributed as part of the standard CDF distribution package
  - See http://nssdc/cdf/html/FAQ.html for a detailed description of each tool

What is CDF

- Self-describing platform-independent scientific data format
- Available on all major platforms
  - Unix (HP, AIX, Sun OS, Solaris, Linux, SGI)
  - Windows 95/98/NT/2000
  - Macintosh OS 8.x, 9.x, 10.x
  - DEC (VMS, OpenVMS, Ultrix, OSF/1)
- Interfaces supported: C, Fortran, Java, and Perl
- See more information at the CDF FAQ page (http://nssdc.gsfc.nasa.gov/cdf/html/FAQ.html)

CDF Markup Language (CDFML)

- XML Markup Language describing Common Data Format (CDF) data and metadata
- A CDFML file (XML file) is an ASCII representation of a CDF file
- XML promotes and facilitates data interoperability with other science data formats (since translation to and from another data format via XSL is easy)

Future Work

- Two-way netCDF-to-CDF translator
- CDF-to-FITS translator
- Sparse array
- Grouping capability
- Ability to handle files that are > 2 GB
- Open to community feedback

CDF Translators

- HDF5-to-CDF
  - Tested against ACE SWICS L2 datasets
- FITS-to-CDF
  - Tested against HEASARC datasets
- netCDF-to-CDF (will be developed in the near future)