

COMMON DATA FORMATS

By Larry Atkinson and Steve Hankin

IMAGINE THAT YOU were handed a disk full of data saying nothing but, "Plot it up!". With common data formats that are machine independent and self-describing, you could do that and expect success.

Imagine you are starting a project that will be a minor, though important, effort over the coming year. You know you will come and go to the project with possible 1-month gaps in your attention to it. The project involves the analysis of various bits of data—some your own, some from others, some extracted from figures in published articles. Some data sets are small; some are large. Some of the data are 1-D data, some are 2-D, some are time series and some are imagery. Unless you are more organized than most of us, we would bet that each time you return to the project you spend a good deal of time recalling what the data are, what the format was, where you got it, and what you intended to do with it. This is when common data formats can help.

If you had put your data into the netCDF format, for example, you would merely type "ncdump filename" to get an immediate look at the data, its attributes, where it came from, etc. From now on we'll refer to netCDF. (See the sources of information on related common data formats at the end of this article.)

We found netCDF useful in the following way. Say you had some old data taken from a figure in a paper and you wanted to keep the source of the data, and how you got it, with the data as you moved it from directory to directory. All you had to do was add the information "Data read by hand from Figure 21 in paper by . . ." to a comment attribute of the file. Then you, or anyone you gave the data to, could check the source, could acknowledge the source, and could check that you extracted the data properly.

So what are common data formats? They are not the same as National Oceanographic Data Center (NODC) data formats. They are a way to move data between different types of machines: they are *device independent*. The data file contains information about itself: it is *self describing*. You will probably never know how the data are stored. You won't know, but every PC, Mac, and Unix machine will. The creators and users of the common data formats provide free tools with which to create, examine, and modify files of data. But it gets better. Most software packages such as IDL, Ferret, etc., provide access tools to the common data formats.

Another example. Assume we have some temperature, salinity, and oxygen data from a station. We can hand you a diskette with one file on it and just say, "here is the data and it is a netCDF file". We don't have to say anything else like "The format is ASCII, 2x, f5.2, 3x, f8.3 . . .". You, with your favorite software package, will open the file, look at the contents and proceed.

Here is an example from the Ferret program (see reference below):

"USE sample.cdf"	Open the file
"SHOW DATA"	Show the variables, say, depth, temp, salt, oxygen
"PLOT temp"	Make a fully labeled plot of temp on its depth axis

and so on.

Imagery is handled as easily as common data. You can even co-mingle imagery and many types of data in one netCDF file. Thus one file could have all of the data relevant to a sampling trip: the CTD data, the great image, the recent weather, and comments on the project.

How do you get started? First see if your analysis software of choice provides

the tools to access CDF's. If it doesn't we suggest switching to one that does! Next create a simple CDF file. That can be done in netCDF by using the Common Data Language (CDL). A CDL file for the example above is included below. Then with the single command "ncgen filename" the netCDF file is created. If that is too complicated, software is provided to merely read columnar data with text headers in.

A wealth of information about NetCDF is available on the Internet. Start with the netCDF home page at

<http://www.unidata.ucar.edu/packages/netcdf>.

This is where you can download the software and documentation. The "Frequently Asked Questions" will guide you to a number of further references both on-line and in the literature.

For some guidelines on setting up your particular data in netCDF look at

<http://www.unidata.ucar.edu/packages/netcdf/conventions.html>.

A list of free and commercial display and analysis packages that work with netCDF is available under

<http://www.unidata.ucar.edu/packages/netcdf/utilities.html>.

The Ferret program shown above is available at

<http://www.pmel.noaa.gov/ferret/home.html>.

Information on the related common data formats is available at

HDF:	http://www.ncsa.uiuc.edu/SDG/Software/HDF
NASA CDF:	http://nssdc.gsfc.nasa.gov/cdf .

Here is the complete CDL file for the example above. The file consists of 4 parts: "dimensions," which are like FORTRAN PARAMETERS that tell the

sizes of arrays; “variables,” which are like FORTRAN DIMENSION statements but they also have “attributes” with extra information about the variables such as the variable’s units; “global attributes,” which contain information about the file as a whole; and “data”, the actual data values. This sample CDL file has salinity, temperature and oxygen data at 16 depths. Normally the latitude/longitude position would also be included in the file (in a manner similar to depth). That has been omitted in the interest of simplicity.

```
netcdf sample {
dimensions:
    DEPTH = 16 ;           // 16 depths
variables:
```

```
    float TEMP (DEPTH) ;
        TEMP:long_name = "Temperature" ;
        TEMP:units = "Deg C" ;
    float SALT (DEPTH) ;
        SALT:long_name = "Salinity" ;
        SALT:units = "psu" ;
    float O2 (DEPTH) ;
        O2:missing_value = -777. ;
        O2:long_name = "Dissolved Oxygen" ;
        O2:units = "ml/l" ;
    float DEPTH (DEPTH) ;
        // the depth axis
        DEPTH:units = "meters" ;
// global attributes:
    :history = "Data collected on Spring
    '95 Explorer cruise" ;
data:
DEPTH = 0, 10, 20, 30, 50, 75, 100,
```

```
    125, 150, 200, 250, 300, 400, 500,
    600, 700 ;
TEMP =
    19.3354, 19.4412, 19.206, 18.7279,
    17.4811, 16.4125, 15.7968, 15.3966,
    15.064, 14.508, 14.0757, 13.5573,
    12.6894, 11.7392, 10.7141, 9.6513 ;
SALT =
    36.0759, 36.083, 36.0781, 36.0665,
    36.0571, 36.0477, 36.0316, 35.9889,
    35.9662, 35.9032, 35.861, 35.7957,
    35.6807, 35.5824, 35.4645, 35.3956 ;
O2 =
    5.4138, 5.4781, 5.5113, 5.5782,
    5.6419, 5.5158, 5.4187, 5.3979,
    5.3354, 5.2644, 5.157, 5.0968,
    4.9171, 4.7374, 4.4964, 4.3405 ;
}
```

□