P1.21 DEVELOPMENT OF GTOOL4 NETCDF CONVENTIONS AND PROCESSOR — SELF DESCRIPTIVE FILE FORMAT FOR MULTI-DIMENSIONAL DATA —

TOYODA Eizi*
Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan
Masaki Ishiwatari
Grad. School for Environmental Earth Sciences, Hokkaido University, Sapporo, Japan
Shin-ichi Takehiro, Kensuke Nakajima
Grad. School of Science, Kyushu University, Fukuoka, Japan
Yoshi-Yuki Hayashi
Grad. School of Science, Hokkaido University, Sapporo, Japan
Project Tem Davis of GFD-Dennou Club

1. INTRODUCTION

Scientific numerical data are growing larger and more complex to behave like storm in which scientists are easily got a lost. Data engineering technology is becoming more and more important to deal with those data storms in many fields of sciences.

It has been common that data are not self-descriptive; scientists are forced to remember how the data are obtained, how the data are processed, and how the figures are plotter from the data. When the amount of data and/or the number of manipulation procedures is large, appropriate management of their information associated with data heavily bothers scientists. Here, metadata is referred to as associated data that describe information associated with the core data. The question is `how metadata should be given?'

The authors have been trying to figure out a needed and sufficient list of metadata through the implementation experiment of `gtool4'. Gtool4 is a collective name of NetCDF conventions, Fortran library, and the toolkit for using them. Gtool4 toolkit is a collection of command-line tools for visualization and simple analyses of arbitrary-dimensional data. The catchword of gtool4 is `*figure from data with one clicking*.'

Gtool4 is named after GTOOL3 (GFD-Dennou Club, 1993). GTOOL3 is a toolkit written in FORTRAN 77, and reads or writes specially designed sequential unformatted file. GTOOL3 has been used for postprocessor of CCSR/NIES AGCM. By the use of NetCDF (Russ et al., 1997), gtool4 is free from the problems of GTOOL3, such as binary format portability problem, byte ordering, or restriction of `lat-lon-height' 3-dimensional space.

* *Corresponding Author Address: TOYODA Eizi, Numerical Prediction Division, Japan Meteorological Agency, Tokyo 100 Japan;*
email: toyoda@npd.kishou.go.jp

2. DESIGN AND IMPLEMENTATION

2.1 NETCDF CONVENTIONS

Design of gtool4 NetCDF conventions consists of two parts: data variable and structured variable. Data variable is normal usage of NetCDF variable. Its attributes are almost compatible with existing NetCDF conventions like COARDS (1995). Structured variable is used for storing graphic objects. The graphic object classes are designed for DCL graphics (GFD-Dennou Club, 1988-2001) as shown in Figure 1. They describe relationship between figure and data efficiently.

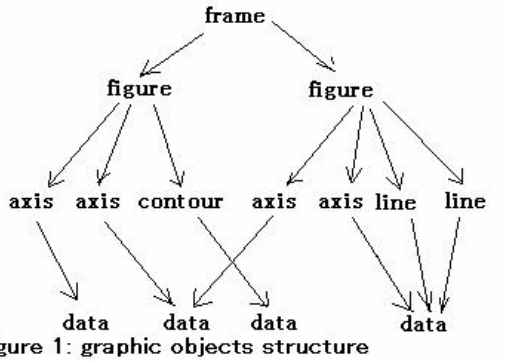

figure 1: graphic objects structure

For compatibility, all new attributes, that are not used commonly among existing major NetCDF conventions, have their name with a common prefix `gt_', so that they can be safely ignored with existing software and conventions. Both data and structured variables can be contained in one NetCDF file, so that it will be `all-in-one' self-descriptive figure that can be reusable as numerical data.

Though all properties of graphic objects are normally stored into structured variables, they do not simplify the initial creation of objects from data variable.

The data variable should `*know*' how it should be rendered. So, the attributes of graphic structured variable can also be attached to the data variables, and they describe `the default graphic behavior' of the data.

2.2 LAYER-STRUCTURED LIBRARY

In order to manage the complexity of software, a layer-structured design is applied (Figure 2). For example, the `gtview' command calls graphic object methods, and they call the generic data access, and they call NetCDF data access, in turn.

Every layer provides the minimized interface using Fortran 90 module. With use of module, users cannot access the private data, and the consistency of internal data structure can be easily assured.

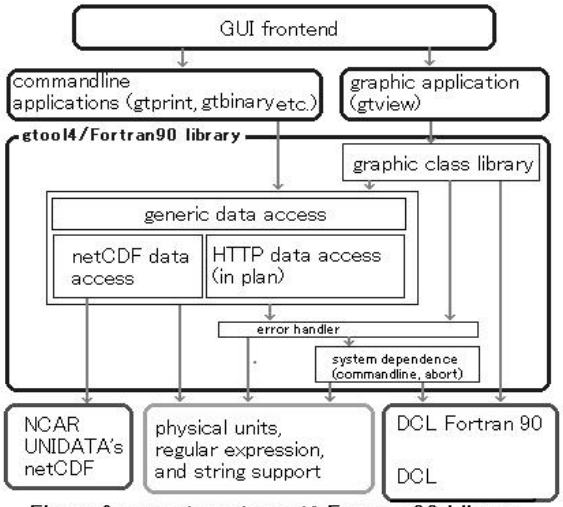

Figure 2: overview of gtool4 Fortran 90 Library and tools

Another benefit of the layer structure is robustness in replacement of adjacent layer. Appropriately designed interface allows changing or complete rewriting the layer without changing user's code at all. Indeed, our last large rewriting was in the generic data access layer, without changing layers above in the figure.

2.3 GENERIC DATA ACCESS

The interface of multi-dimensional array data is more simplified than that of NetCDF.

Essentially, the input/output interface is modeled on that of ordinal file access (OPEN, GET, PUT, CLOSE). Unlike NetCDF, the user doesn't need to manage separate handles for dataset and variable. OPEN is called with variable handle and variable locator string like `file?var' notation, and an internal dataset handle is managed automatically. The variable locator is conformant with the notation of URL (Berners-Lee, 1994).

Normally, GET or PUT provides an access to the whole array. Access to the partial variable can be done with the same interface after calling LIMIT. Unlike NetCDF, no additional arguments needed for GET or PUT.

2.4 TOOLS

Commands with simple function (so called `one liner') like addition, subtraction, or multiplication, are provided. Their code is quite simple, because the GET/PUT interface described above is independent of the number of dimensions. Likewise, the general viewer `gtview' can draw contours or line graph from any slice of multi-dimensional array data.

Commands for more complex calculation like FFT, derivation, or integration are now in progress.

3. CONCLUDING REMARK

To cope with a distributed environment, demand to direct network access is increasing. We are now planning to support HTTP data access. Using network access, automatic conversion from non-NetCDF formats or aggregating distributed datasets into uniform array will be easily implemented.

All gtool4 software and documentation are available at http://www.gfd-dennou.org/arch/gtool4/.

REFERENCES

Berners-Lee, T. et al., 1994: Uniform Resource Locators (URL). RFC 1738. ftp://ftp.isi.edu/in-notes/rfc1738.txt

Cooperative Ocean/Atmosphere Research Data Service, 1995: Conventions for the standardization of NetCDF files. http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html

GFD-Dennou Club, 1988-2001: Dennou Club Library (DCL). http://www.gfd-dennou.org/arch/dcl/

GFD-Dennou Club, 1993: GTOOL3. http://www.gfd-dennou.org/arch/gtool/SIGEN.htm

Russ et al., 1997: NetCDF User's Guide For Fortran, version 3. http://www.unidata.ucar.edu/packages/netcdf/guidef/