

多次元数値データの自己記述的格納形式 gtool4 の開発

Development of gtool4: a self-descriptive storage format for multi-dimensional numerical data

豊田 英司[1], 石渡 正樹[2], 林 祥介[3], 堀之内 武[4], 赤堀 浩司[5], 沼口 敦[2], 地球流体電脳倶楽部 Davis Project 林 祥介

Eizi TOYODA[1], Masaki Ishiwatari[2], Yoshi-Yuki Hayashi[3], Takeshi Horinouchi[4], Koji Akahori[5], Atusi NUMAGUTI[6], GFD Dennou Club Davis Project Hayashi Yoshi-Yuki

[1] 東大数理, [2] 北大・地球環境, [3] 北大・理・地球惑星, [4] 京大・超高層, [5] 名大・工・計算理工
[1] Math. Sci., Univ. Tokyo, [2] Graduate School of Environmental Earth Science, Hokkaido University, [3] Earth and Planetary Sci., Hokkaido Univ., [4] Radio Atmos. Sci. Center, Kyoto Univ., [5] Computational Sci. and Eng., Nagoya Univ., [6] EES, Hokkaido Univ.

<http://www.gfd-dennou.org/arch/davis/>

地球流体现象の研究での利用を想定した、多次元数値データ格納のためのファイル形式を開発した。機種依存性をなくすため低レベルでは netCDF を採用し、その netCDF 規約として、自己記述性、格子構造に限らない自由なデータ構造、可視化情報の並存を特徴とする gtool4 規約 をあらたにまとめた。自己記述性の検証のため Fortran 90 で処理系を作成している。

1. はじめに

観測手段や計算機の発達により、地球流体现象の研究で用いる数値データは大規模化している。このため数値データの処理技術が研究の効率や信頼性を大きく左右するようになってきている。

従来地球流体科学コミュニティで用いられてきた数値データ処理環境ではデータは自己記述的ではなく、数値処理や描画に際して作業者がデータに付随する情報を与えることを前提としていた。大量のデータ処理においては付随情報の適切な管理は大きな負担になる。

数値データの付随情報は関係データベースによって管理できるように思われる。しかし地球流体分野で用いられるデータは大規模な多次元数値データを主とすることに特徴がある[2]。また、管理すべき情報のリストは標準化の途上にあり、特に、可視化に関する情報についてはほとんど議論がなされていない。

そこで本研究では可視化を意識した多次元数値データの自己記述的な格納形式と処理系の開発を試みた。標語は「クリックするだけで適当な絵が描けるデータファイル」である。

2. 設計方針

本研究では以下のような要件を満たすようなデータの格納形式の設計を行った。

自己記述性

数値データに与えるべき付随情報はすべて同じファイルにまとめて格納できるべきである。これによって付随情報の散逸が防止され、データの流通性が向上する。

格子状構造に限定しないこと

非格子状構造データも格子状データと統一的に取り扱うことができねばならない。多数の地上気象観測点や漂流ブイでの観測結果のような非格子状データを同時に扱うことを目標としているからである。

数値データと可視化情報の統合

数値データと可視化情報（たとえばグラフの軸の取り方）をひとつのファイルに格納できるべきである。これ

により、数値データを可視化するための情報の散逸を防止することができる。

プラットフォーム非依存性

データをファイルに格納する際にはプラットフォーム依存しないデータ表現法を用いるべきである。これにより、作業者は浮動小数点数形式、バイトオーダー、文字コードなどの異なる環境にファイルを転送してもそれらの違いを意識する必要がない。

3. gtool4 netCDF 規約

上述の設計方針をふまえて、本研究では低レベルファイル形式として netCDF[2] を用いたデータの格納形式を考察した。NetCDF は多次元数値データとそれに付随する情報を格納する機構を提供する。

この機構を活用することにより、自己記述的な数値データの表現と可視化情報の統合を行った gtool4 netCDF 規約 <URL:<http://www.gfd-dennou.org/arch/gtool4/conventions/>> を作成した。gtool4 規約は COARDS 規約[3] および NCAR CSM 規約[4] の上位互換性を意識して設計されている。

5. Fortran 90 による処理系実装実験

gtool4 規約に従う netCDF ファイルの処理系の実装実験 <URL:<http://www.gfd-dennou.org/arch/gtool4/current/>> を行った。プログラムは Fortran 90 で記述している。ISO/IEC 1539-2:1994 に準拠した可変長文字列ライブラリが用意されており、内部で使用される文字列には長さ制限がない。

謝辞

本研究は 科学技術振興事業団 の計算科学技術活用型特定研究開発推進事業(短期集中型)「地球惑星流体現象を念頭においた多次元数値データの構造化」の一環として行われた。

世界における netCDF の利用状況に関して教示していただいた Dr. Byron Bouville (NCAR) に感謝する。

参照

- 1) 地球流体電脳倶楽部 davis プロジェクト: <URL:<http://www.gfd-dennou.org/arch/davis/>>, <URL:mailto:dennou_davis@gfd-dennou.org>
- 2) netCDF: <URL:<http://www.unidata.ucar.edu/packages/netcdf/>>
- 3) COARDS conventions: <URL:http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html>
- 4) NCAR CSM conventions: <URL:<http://www.cgd.ucar.edu/cms/eaton/netcdf/NCAR-CSM.html>>