

地球流体解析のための数値データの構造化

北海道大学 理学研究科 地球惑星科学専攻 林 祥介 (shosuke@gfd-dennou.org)

【はじめに】

地球・惑星での流体现象(大気や海洋、マントルや中心核)の研究・教育における困難は、観測や数値計算によって得られた多次元データ(多くは時空間場のデータ)を認識可能な形に整形するのに多大な労力がかかることにある。これらのデータは、通常、大規模(サイズが大きい)か、または、サイズは小さくなくとも類似データが大量にある)であり、かつ、厄介なことに多様である。グループやプロジェクトが異なれば、データの構造が異なり、データ毎に個別の道具を用意するのが常である。数値実験のデータと衛星観測データとは容易に結合されねばならないし、別々なグループで行なわれた実験観測の比較検討を進める際にも異なるデータの同等な加工表示が必要になるのだが、現状は著しく効率が悪く、研究者や学生は容易に「データに埋もれ」てしまう。

流体现象の理解も数式による認識だけでなく絵による認識が重要となりつつある。しかしながら、通常の紙媒体上での記述は多次元データの特定の断面(注目すべき特徴)を表示したものに過ぎない。昨今電子化されつつある記述においても紙媒体での習慣が踏襲されており、データから自分の好きな断面を自分で再構成し理解を深めることは不可能である。しかるに、理解を深めるためには、数式を加工し認識を共有するのと同様のことを多次元データにおいてもたやすく実現したいのである。従来の教科書がデータの特定の断面しか表示しえなかったのに対し、これからの電子化された教科書では、標準的な多次元データをバイナリ形式で内在し、必要に応じて違う断面や違う可視化を行うことができ、あるいは、理論計算のためのデータが生成できるようになっていなければならないと考えているわけである。

このような状況、将来像に対して我々が最終的にめざすべき目標は、流通加工が容易となるような標準的なデータ構造をもつことである。これは、「クリック一発で絵が書ける数値データ」を整えることに他ならない。とりあえず製作者の意図した断面を表示してくれて、しかも、必要に応じて加工生成流転できるものでなければならない。また、研究教育現場に耐えられるフリー(あるいは十分に廉価)な環境によっても加工表示が実現可能でなければならない。かつ、手元の計算情報環境から分散計算環境(リモートスーパーコンピュータ)までを想定した何らかの意味でネットワーク透過なものでなければならない。

本開発研究においては、上記認識にもとづき、多次元データに対する扱いを

- ・データの構造
- ・データの処理
- ・データの可視化

の三つの層に切り分けて考察し、これらの各層に対して、

- ・Fortran 90 または C を用いた、既存パラダイム上での数値計算エンジンの整備
- ・オブジェクト指向言語 Ruby、新しいパラダイムでの試み

の二つの方向から実装を試みた。

【データの構造: gtool4 規約】

標準化されるべきデータ構造を提案するにあたって、最初に、そして、くりかえし検討されるべき課題は、地球惑星現象に関連したデータが自己記述するべき属性リストの作成である。これは、研究教育においてデータをどのように加工・可視化しているかという我々の日常作業の定型化である。属性リストは、具体的なソフトウェアや計算機上でのデータ表現方法の詳細には依存しない、抽象データとして検討することが望ましい。しかし、ここでは具体的実装に直結すべく、データ格納形式の一つである NetCDF 形式上に具体的な実装規約をつくることでデータ構造の検討を行うことにした。

NetCDF (network Common Data Form) は多次元配列型のデータを格納する形式とそれを操作するためのインターフェースソフトウェア群である。インターフェースが実装された計算機間においては NetCDF 形式バイナリデータをネットワーク等により転送し直接読みだすことができる。NetCDF 形式はファイルの中に任意個の変数(多次元の配列)を定義

することと、ファイルや変数に任意個の属性を付与することをサポートする変数名や「NetCDF ユーザーズガイド」で規定された若干のもの以外の属性名の意味や用法は NetCDF の利用者の自由であり、利用者が規約を定めることになっている。ここでは、我々の過去の資源である GTOOL3 での経験と NetCDF 形式上での既存規約である GDT、COARDS、CSM を参照して、我々目的に応じた規約 (NetCDFgtool4 規約) の構築実験を行った。

gtool4 規約は既存規約の上位互換としながら、特に可視化に関する属性の補強を行った。これにより、大規模データをもとにした作画作業を効率化を可能にするとともに、可視化情報の伝達においても大きな効果を発揮すると期待される。

【Fortran 90 による実装: gtool4】

gtool4 は、GTOOL3 を後継するべく、NetCDF gtool4 規約に従ったデータを処理(生成加工)するソフトウェア群である。言語は Fortran 90 を使っている。データの可視化に関しては、NetCDF gtool4 規約にしたがったデータに対し DCL Fortran 90 インターフェースを通して可視化することにし、教育資源としても耐えられるフリーな環境提供を保証している。既存の商用可視化ソフトウェアである AVS や IDL を用いても、そのスクリプトソフトウェア群を作成し NetCDF gtool4 規約を解釈させれば可視化可能であるが、現時点では必要ユーティリーの整備は行っていない。

【Ruby での試み】

将来への試みとして、国産オブジェクト指向スクリプト言語 Ruby による実装実験を行った。データ構造は NetCDF gtool4 規約を Ruby スクリプトとして格納するものである。我々は、Ruby スクリプトファイルから NetCDF 形式ファイルを入出力するためのインターフェースルーチンを用意した。NetCDF 形式ファイルを用いることによりネットワーク透過なデータ転送が可能となる。また、データの処理を Ruby 上の演算で行えるようにするべく、多次元数値配列を操作するためのクラスライブラリを作成した。データ可視化に関して、Ruby から DCL C インターフェースを呼び出す方式についての検討も行われた。

【可視化系の整理・固定: DCL-5.1】

教育環境でも安心して利用提供可能なフリーな可視化道具の保証を行うべく、我々がこれまで作成してきたライブラリである DCL を整備固定し、マニュアルの整備と Fortran 90 インターフェースの作成、C 化への手順の確立を行った。現在 DCL-5.1 として公開している。

【今後の展望】

データ構造に関しては、現時点では、観測データのように不規則格子データ・ベクトル的な場・座標表現への対応は不十分である。また、可視化を想定した属性は2次元の作図のみを考慮したものでしかなく3次元化が急務である。データの処理に関しては、FORTRAN90 においても Ruby においても我々が必要としている諸演算がまだ適宜実装されていない。今後の課題として残されている。

Ruby を用いた実装における現状の問題点は Ruby コミュニティーにおいて数値処理の経験が少ないことにある。スーパーコンピュータ上への効率良い実装、巨大データに対する実装も問題となっている。同種のスクリプト言語に Python があるが、Ruby のほうが設計思想がわかりやすく、かつ、開発者グループと日本語でコミュニケーションできることがもっとも大きなメリットである。本開発研究活動に刺激され Ruby の math グループが立ち上がり、多次元配列を扱う ruby の仕様決定が進められつつある。このグループの活動に期待したい。

NetCDF は米国においてその下部構造を HDF に統合し並列化対応することが計画されている。今後並列計算機とのデータの交換が盛んになると思われるのでこの方向の展開も期待される。

【参考 URL】

COARDS http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html
CSM <http://www.cgd.ucar.edu/cms/eaton/netcdf/NCAR-CSM.html>
DCL <http://www.gfd-dennou.org/arch/dcl>
HDF <http://hdf.ncsa.uiuc.edu/>
GDT http://www.pcmdi.llnl.gov/drach/GDT_convention.html
GTOOL3 <http://www.gfd-dennou.org/arch/gtool>
Ruby <http://www.ruby-lang.org/>

なお、本開発研究でのプロダクトは地球流体電脳倶楽部サーバー上の公開領域
<http://www.gfd-dennou.org/arch/davis/>
から適宜たどれるようになっている。

これは平成12年3月9日に開催した
計算科学技術活用型特定研究開発推進事業
研究報告会（主催 科学技術振興事業団）
の予稿集から抜粋したものです。