

すべての地球学データに使えるデータベースの要件- gtool4 プロジェクトを通じて見えてきたもの -

Requirement for comprehensive database for geosciences --- current product and prospect of gtool4 project

豊田 英司[1], 石渡 正樹[2], 竹広 真一[3], 林 祥介[4], 地球流体電脳倶楽部 Davis Project 林 祥介

Eizi TOYODA[1], Masaki Ishiwatari[2], Shin-ichi Takehiro[3], Yoshi-Yuki Hayashi[4], GFD Dennou Club Davis Project Hayashi Yoshi-Yuki

[1] 気象庁 数値, [2] 北大・地球環境, [3] 九大・理・地惑, [4] 北大・理・地球惑星

[1] JMA NPD, [2] Graduate School of Environmental Earth Science, Hokkaido University, [3] Earth and Planetary Sci., Kyushu Univ., [4] Earth and Planetary Sci., Hokkaido Univ.

<http://www.gfd-dennou.org/arch/gtool4/>

本大会講演 (2000 年、2001 年) でも報告してきたように、著者らは大規模な地球科学数値データを念頭においた解析可視化環境 gtool4 の開発を行ってきた。作業の進展によって最低限の実装目標、すなわち沼口敦氏による GTOOL3 の機能の実現と gtool4 netCDF 規約の実証にはほぼ目処がついてきた。そこで gtool4 の成果を報告するとともに、プロジェクト本来の目的に立ち返り、地球惑星科学において暗黙理にデータに期待されている概念的構造(データモデル) に関して本実験を通じてわかったことをまとめておきたい。

GTOOL3 の継承・発展

GTOOL3 は 3 次元規則格子を前提として、データの切り出し、四則演算、軸方向平均、作図等の機能を提供していた。Gtool4 では netCDF ファイル形式の上で同様のライブラリとツールを再構成し、ファイルの可搬性を確保し、任意の属性情報を付加可能にし、次元数に関する制約を除去している。

入出力インターフェイスは Fortran 90 の構造型と総称名を活用してデータの次元数に依存しないように抽象化されている。特に数値モデルでの利用を想定し、時系列の書き出しに特化して簡潔化したインターフェイスも用意されている。これによってたとえば (x, y, z, t) 4 次元の時系列から (y, z) 2 次元の東西平均年間場を差し引くといった次元構成の異なるデータの対に関する要素別処理が簡潔に記述できるようになった。この処理は実用上よく用いられるにもかかわらず GTOOL3 では次元毎の個別対応が必要で実現困難だった。

Gtool4 netCDF 規約の実証

データ形式を規定する gtool4 netCDF 規約はプログラムの実装前に策定した。このため、実装上発生する問題を通じて当初の明示的考察とデータ処理における

暗黙理の期待との乖離が明らかになってきた。今のところ以下のような問題が認識されている。

まず、可視化情報を管理するためのクラス設計に修正が加えられた。当初 DCL のパッケージと一対一対応するようにクラスを設けたが、実装の過程でデータ依存性が類似するクラスを統合したほうがプログラムの見とおしができることがわかった。等値線と塗り分け、線グラフと散布図が統合されている。これは Fortran 90 に継承がないためである。

ついで、規約の扱う netCDF 属性について、データ加工ツールによって作成されるデータに自動的に適切な属性を付与するための規則の必要性が明らかになった。最低限作業ができるための必須性はよく議論されていたが、特に複数の入力によって生成されるファイルについての規則は不十分だった。多くのアプリケーションで統一的な扱いをするためには属性付与方式を共通化し netCDF 変数作成のための標準インターフェイスとして実装すべきだろう。

データモデルに関する展望

GTOOL3 と同様、現在の gtool4 の実装では、netCDF が提供する多次元配列を用いて、配列添字が表示位置と直接対応する可視化だけを行っている。しかし地球惑星科学一般のデータはこのような規則格子データモデルとは限らない。例えば地点観測データの識別は地点番号が用いられるが、地点番号は第一義的には名前だけに対応し位置とみなせない。図化に必要な経緯度情報は地点番号をキーとして探索することになる。このようなデータは gtool4 netCDF 規約策定当初から考慮されているが、その実装はまだ手付かずである。ここではより一般的に

データモデルという枠組からとらえなおしてみたい。3つのデータモデルが議論された：規則格子、関係モデル、木構造である。Gtool4 規約は変数の扱いで規則格子と関係モデルを、ファイル名を用いたデータ指示で木構造を扱う。データモデルは効率の犠牲のもとで相互に変換できる。一方で、gtool4 は規則格子パラダイムを超えることができなかつた。つまり、ソフトウェアはデータよりもずっと強くデータモデルに依存している。

したがって、今後期待される展開としては、さまざまなデータモデルに対応した万能ソフトウェアを目指すより、データモデル依存のツールが多様なデータを自分の都合がいいように読みとることができるようにする、というほうが生産的だろう。

クライアント・サーバシステムはこの問題への解となりうる。たとえば配列構造 表構造 木構造の写像を策定すれば、同じ HTTP/URL 資源を apache でキャッシュし SQL で検索し格子状に扱えるところだけ gtool4 で演算・可視化する、といったことができるようになるに違いない。また一つの実体から多様なファイル形式を提供することも容易になる。このようになってはじめてすべての地球科学データを統合できるようなデータベースを作れるようになるであろう。

参照

Swamp project, GTOOL3.

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

GFD-Dennou Club Davis Project, Gtool4.

<http://www.gfd-dennou.org/arch/gtool4/>